Thérèse Phan
Jean-Pierre Rowenczyk

Exercices et problèmes de statistique et probabilités

2e édition

DUNOD

Illustration de couverture : *digitalvision*

ISBN 978-2-10-056298-5

Table des matières

Avertissement

Cet ouvrage est destiné aux étudiants de Licence, de première année des Grandes Écoles d'ingénieurs, de commerce et de gestion ou d'Institut Universitaires de Technologie désireux d'appréhender les concepts et les notions de base de la statistique.
Il peut être utile à tous ceux qui seraient désireux d'acquérir ou de revoir les notions opérationnelles des méthodes de base de la statistique.

Cet ouvrage comporte des rappels de cours sans démonstrations, des exercices classiques de difficultés progressives (le niveau de difficulté est repéré par un nombre d'étoiles), ainsi que des problèmes plus complexes permettant d'aborder des cas concrets d'utilisation de la statistique dans différents domaines d'application. Il est découpé en chapitres mais il comporte fondamentalement deux grandes parties :

- Une première partie concerne le calcul des probabilités
  Bien que comportant des rappels de cours relativement complets, nous avons choisi, délibérément, de ne proposer dans cette partie, que des exercices abordant des notions et des calculs de probabilité qui sont utilisés en statistique : Théorème Central-Limite (ou théorème de la limite centrale), Lois de probabilités fréquemment utilisées en statistique (Loi normale, du Khi-deux, de Student, de Fisher...)
  Nous avons donc évité de proposer des exercices de probabilités calculatoires classiques (exercices utilisant la combinatoire, calcul de paramètres de lois de probabilités...).
  Pour cette raison, avant d'aborder les chapitres de statistique, nous conseillons vivement au lecteur, de se reporter, en cas de besoin, aux ouvrages spécialisés, afin de revoir ou de compléter leurs connaissances en matière de calcul des probabilités.
- Une deuxième partie est consacrée à l'étude des trois méthodes de base utilisées en statistique :
  - L'estimation ponctuelle
  - L'estimation par intervalle
  - Les tests d'hypothèse

  Les chapitres concernant l'estimation ponctuelle permette d'aborder les notions essentielles permettant d'étudier les estimateurs de paramètres réels de lois de probabilités.
  Néanmoins, ces chapitres proposent quelques exemples d'estimation de paramètres vectoriels.
  Les chapitres consacrés à l'estimation par intervalle proposent un éventail large d'exercices différents, permettant d'appréhender la plupart des cas concrets rencontrés dans les différents domaines utilisant la statistique.
  Les chapitres consacrés aux tests d'hypothèses sont essentiellement consacrés à l'étude des tests paramétriques dans le cas d'hypothèses simples et à l'étude de deux types de tests non paramétriques, les tests d'ajustement et les tests d'indépendance.

Les différents chapitres proposent toujours la même organisation : les énoncés, puis une rubrique « Du mal à démarrer », et enfin, les corrigés des exercices proposés.
Chaque corrigé propose, en outre, un bilan « ce qu'il faut retenir ».

Remerciements

Nous tenons, tout d’abord à exprimer toute notre gratitude à nos collègues de l’École Centrale de Paris et de l’École Spéciale des Travaux Publics, pour nous avoir incités à élaborer cet ouvrage et pour nous avoir fourni de nombreux conseils de rédaction.
En particulier, nous tenons à remercier, Alain MARRET et Michel LUCIEN, pour leur apport lors de l’élaboration du contenu de cet ouvrage.

Nos remerciements vont ensuite à Franck PHAN, pour son aide précieuse pour l’utilisation de Latex et donc de la réalisation de la maquette de cet ouvrage.

Enfin, nous tenons également à remercier vivement les Éditions DUNOD, Anne Bourguignon et Benjamin Peylet, pour leur accueil, leur compétence et leur grande compréhension au cours de la réalisation de cet ouvrage.

Thérèse PHAN et **Jean-Pierre ROWENCZYK**

1

Probabilités

RAPPEL DE COURS

1.1 Rappels de Mathématiques

a) Opérations sur les ensembles

Soit Ω un ensemble et A, B ... des parties de Ω. Si $\overline{A}$ désigne le complémentaire de A dans Ω, alors nous avons :

- $A \cup \overline{A} = \Omega$
- $\overline{A \cup B} = \overline{A} \cap \overline{B}$
- $A \cup B = (A \cap \overline{B}) \cup (A \cap B) \cup (\overline{A} \cap B)$
- $A \equiv \cup_i (A \cap B_i)$
  - ▷ si les évènements B_i sont incompatibles entre eux
  - ▷ et si $\Omega \equiv \cup_i B_i$

b) Analyse combinatoire

Nous rappellons ici quelques résultats :

- Nombre d'arrangements de p objets pris parmi n avec répétition

$$\Re_n^p = n^p$$

- Nombre d'arrangements de p objets pris parmi n sans répétition

$$A_n^p = n(n-1)\ldots(n-p+1) = \frac{n!}{(n-p)!}$$

- Nombre de combinaisons de p objets pris parmi n avec répétition

$$K_n^p = C_{n+p-1}^p$$

- Nombre de combinaisons de p objets pris parmi n sans répétition

$$C_n^p = \frac{n!}{p!(n-p)!} = \frac{A_n^p}{p!}$$

- Nombre de permutations de n objets

$$Per(n) = n!$$

1.2 Axiomes du calcul des probabilités

a) Généralités

La théorie des probabilités repose sur l'étude de phénomènes aléatoires. Une expérience est dite aléatoire si on ne peut pas prévoir son résultat et si répétée dans les mêmes conditions, elle peut donner des résultats différents. Les résultats possibles de cette expérience constituent l'ensemble fondamental Ω. Un événement aléatoire est une assertion relative au résultat de l'expérience. On identifie usuellement l'événement aléatoire et la partie de Ω pour laquelle cet événement est réalisé.

Si P est une probabilité définie sur Ω, et si A et B sont deux parties de Ω, on a :

- $P(\varnothing) = 0 \quad \text{et} \quad P(\Omega) = 1$
- $P(\overline{A}) = 1 - P(A)$
- $P(A \cup B) = P(A) + P(B) - P(A \cap B)$



$P(A \cup B) = P(A) + P(B) \quad \text{si} \quad A \cap B = \varnothing$

b) Probabilités conditionnelles

On définit la probabilité conditionnelle de l'événement A sachant que l'événement B est réalisé, par :

$$P(A/B) = \frac{P(A \cap B)}{P(B)}$$

c) Formule de décomposition

Si l'ensemble des parties U_j de Ω forme un système complet d'événements, c'est-à-dire si les U_j sont indépendants et si leur réunion forme Ω tout entier, alors :

$$P(A) = \sum_{j=1}^{n} P(A/U_j)P(U_j)$$

d) Indépendance de deux événements

$$A \text{ et } B \text{ indépendants} \quad \Leftrightarrow \quad P(A \cap B) = P(A)P(B)$$

e) Probabilités des causes ou probabilités de BAYES

$$P(A/B) = \frac{P(A \cap B)}{P(B)} = \frac{P(B/A)P(A)}{P(B)}$$

Si l'ensemble des parties A_i de Ω forme un système complet d'événements,

$$P(A_k/B) = \frac{P(B/A_k)P(A_k)}{\sum_i P(B/A_i)P(A_i)}$$

1.3 Notion de variable aléatoire

Lorsque l'ensemble fondamental Ω est tout ou partie de l'ensemble des réels R, le concept d'événement aléatoire est remplacé par celui de variable aléatoire. On distingue usuellement :

1. les variables aléatoires discrètes pour lesquelles l'ensemble Ω est un ensemble discret de valeurs numériques (par exemple N ensemble des entiers naturels)
2. les variables aléatoires continues pour lesquelles l'ensemble Ω est un intervalle de R ou R tout entier.

a) Fonction de répartition

On appelle « Fonction de répartition d'une variable aléatoire X » l'application F de R dans $[0, 1]$ définie par :

$$F(x) = P(X < x)$$

b) Variable aléatoire discrète

▷ On définit la probabilité attachée en un point x du domaine de définition de la variable aléatoire X discrète par :

$$P(X = x)$$

▷ Fonction de répartition de X :

$$F(x) = P(X < x) = \sum_{t<x} P(X = t)$$

c) Variable aléatoire continue

▷ On dit que la variable aléatoire X de fonction de répartition F est continue si on peut définir une fonction densité de probabilité f de X vérifiant :

$$f(x) = F'(x) \quad \text{ou} \quad F(x) = \int_{-\infty}^{x} f(t)dt$$

▷ La probabilité attachée au segment $[a, b]$ est alors :

$$P[a \leqslant X \leqslant b] = \int_{a}^{b} f(x)dx = F(b) - F(a)$$

d) Formule de changement de variables

- Cas discret

$$p_y = P(Y = y) = P(X \in \varphi^{-1}(y)) = \sum_{x\in\varphi^{-1}(y)} P(X = x)$$

- Cas continu
  - ▷ φ est monotone croissante

$$G(y) = P(Y < y) = P(X < \varphi^{-1}(y)) = F\left[\varphi^{-1}(y)\right]$$

  - ▷ φ est monotone décroissante

$$G(y) = P(Y < y) = P(X \geqslant \varphi^{-1}(y)) = 1 - P[X < \varphi^{-1}(y)]$$
$$G(y) = 1 - F\left[\varphi^{-1}(y)\right]$$

  - ▷ φ n'est pas monotone

$$G(y) = P(Y < y) = P(X \in I_1) + \cdots + P(X \in I_n)$$

où $I_1, \ldots, I_n$ sont les intervalles de la variable aléatoire X qui correspondent au domaine $Y < y$.

e) Fonction caractéristique φ_X

- Cas discret

$$\varphi_X(t) = E\left(e^{itX}\right) = \sum_{D_X} \exp(itx)P(X = x)$$

- Cas continu

$$\varphi_X(t) = E\left(e^{itX}\right) = \int_{D_X} \exp(itx)f(x)dx$$

- Exemples de fonctions caractéristiques :
  - ▷ Loi binomiale $B(n, p)$ $\quad \varphi_X(t) = (pe^{it} + 1 - p)^n$
  - ▷ Loi de Poisson $P(\lambda)$ $\quad \varphi_X(t) = \exp\left[\lambda(e^{it} - 1)\right]$
  - ▷ Loi de Gauss $LG(m, \sigma)$ $\quad \varphi_X(t) = e^{itm} \times \exp\left(\frac{-t^2\sigma^2}{2}\right)$

f) Fonctions génératrice G

La fonction génératrice des moments de la variable X est définie par :

$$G_X(u) = E(e^{uX})$$

1.4 Moments d'une variable aléatoire

a) Moment d'ordre r par rapport à l'origine

➤ Calcul direct

- ▷ variable discrète $\quad m_r = E(X^r) = \sum_X x_i^r P(X = x_i)$
- ▷ variable continue $\quad m_r = E(X^r) = \int_{D_X} t^r f(t)dt$

➤ Utilisation de la fonction caractéristique

$$m_n = E(X^n) = \frac{\varphi_X^n(0)}{i^n}$$

➤ Utilisation de la fonction génératrice (pour les variables discrètes)

$$E\left[X(X-1)\ldots(X-n+1)\right] = G_X^{(n)}(0)$$

où $G_X^{(n)}$ est la dérivée d'ordre n de la fonction génératrice.

b) Moments centrés d'ordre r

▷ variable discrète

$$\mu_r = E\left[(X - E[X])^r\right] = \sum_i (x_i - E[X])^r P(X = x_i)$$

▷ variable continue

$$\mu_r = E\left[(X - E[X])^r\right] = \int_{D_X} (t - E[X])^r f(t)dt$$

c) Espérance mathématique

➤ Calcul direct

▷ variable discrète

$$E(X) = \sum_X x_i P(X = x_i)$$

▷ variable continue

$$E(X) = \int_{D_{X_x}} t f(t)dt$$

➤ Espérance d'une somme

$$E(X+Y) = E(X) + E(Y)$$

➤ Utilisation de la fonction génératrice pour une variable aléatoire discrète

$$E(X) = G_X^{'}(1) = \sum_{D_X} x P(X = x)$$

➤ Utilisation de la fonction caractéristique

$$m_1 = E(X) = \frac{\varphi_X^1(0)}{i^1}$$

- Espérance mathématique de $Y = \varphi(X)$
  - ▹ Variable discrète

$$E[\varphi(X)] = \sum_{D} \varphi(x) P(X = x))$$

  - ▹ Variable absolument continue

$$E[\varphi(X)] = \int_{D_X} \varphi(x) f(x) dx$$

d) Variance

- Calcul direct
  - ▹ Variable discrète

$$\mu_2 = E\left[(X - E[X])^2\right] = \sum_{X} (x_i - E[X])^2 P(X = x_i)$$

  - ▹ Variable continue

$$\mu_2 = E\left[(X - E[X])^2\right] = \int_{D_X} (t - E[X])^2 f(t) dt$$

- Formule développée de la variance

$$V(X) = E(X^2) - [E(X)]^2 = m_2 - m_1^2$$

- Écart-type

$$\sigma_X = \sqrt{V(X)}$$

- Covariance de deux variables X et Y

La covariance des deux variables X et Y est définie par :

$$\text{Cov}\,(X, Y) = E\,[(X - E(X))\,(Y - E(Y))]$$

- Variance d'une somme

$$V(X + Y) = V(X) + V(Y) + 2\text{Cov}\,(X, Y)$$

- Inégalité de Bienaymé-Tchebichev

Soit X une variable aléatoire de moyenne m et d'écart-type σ, alors pour tout t :

$$P(|X - m| > t\sigma) \leqslant \frac{1}{t^2}$$

1.5 Variables à deux dimensions

a) Définitions

➤ Variables discrètes

▷ Probabilité en un point du domaine

$$p_{ij} = P(X = x_i, Y = y_j)$$

▷ Fonction de répartition

$$F(x, y) = P(X < x, Y < y) = \sum_{u<x} \sum_{v<y} P(X = u, Y = v)$$

➤ Variables continues

▷ Fonction de répartition

$$F(x, y) = P(X < x, Y < y) = \iint_{D_{XY}} f(x, y)dxdy$$

▷ Densité de probabilité

$$f(x, y) = \frac{\partial^2 F(x, y)}{\partial x \partial y}$$

Dans ce qui suit, les formules sont données pour des variables continues. Les formules, pour les variables discrètes, s'en déduisent aisément.

b) Lois marginales et lois conditionnelles

➤ Loi marginale de X

$$F(x, .) = \int_{-\infty}^{x} \left(\int_{-\infty}^{+\infty} f(u, v)dv \right) du$$

$$\frac{dF(x, .)}{dx} = f(x, .) = \int_{-\infty}^{+\infty} f(x, v)dv$$

➤ Loi marginale de Y

$$F(., y) = \int_{-\infty}^{y} \left(\int_{-\infty}^{+\infty} f(u, v)du \right) dv$$

$$\frac{dF(., y)}{dy} = f(., y) = \int_{-\infty}^{+\infty} f(u, y)du$$

➤ Loi conditionnelle de Y/X

$$F(y/X = x) = F(y/x) = \frac{\dfrac{\partial F(x,y)}{\partial x}}{\dfrac{dF(x,.)}{dx}}$$

$$f(y/x) = \frac{dF(y/x)}{dy} = \frac{f(x,y)}{f(x,.)}$$

➤ Loi conditionnelle de X/Y

$$F(x/Y = y) = F(x/y) = \frac{\dfrac{\partial F(x,y)}{\partial y}}{\dfrac{dF(.,y)}{dy}}$$

$$f(x/y) = \frac{dF(x/y)}{dx} = \frac{f(x,y)}{f(.,y)}$$

c) Espérance mathématique

➤ Espérances des variables marginales

$$E(X) = \iint_{D_{XY}} xf(x,y)dxdy \qquad E(Y) = \iint_{D_{XY}} yf(x,y)dxdy$$

➤ Espérances des variables conditionnelles

$$E(X/Y) = \int_{X_y} xf(x/y)dx \qquad E(Y/X) = \int_{Y_x} yf(y/x)dy$$

On remarquera que ces espérances sont elles-mêmes des variables aléatoires.

➤ Théorème de l'espérance totale

$$E(X) = E[E(X/Y)] \qquad E(Y) = E[E(Y/X)]$$

d) Variance

➤ Variances conditionnelles

$$V(X/Y) = E[X - E(X/Y)]^2 \qquad V(Y/X) = E[Y - E(Y/X)]^2$$

➤ Théorème de la variance totale

$$V(Y) = V[E(Y/X)] + E[V(Y/X)]$$

e) Corrélation

➤ Coefficient de corrélation

$$\rho = \frac{E\left[(X - E(X))(Y - E(Y))\right]}{\sqrt{V(X) \times V(Y)}} = \frac{\text{Cov}\,(x, y)}{\sigma_x \sigma_y}$$

➤ Rapport de corrélation

$$\eta^2_{Y/X} = \frac{V[E(Y/X)]}{V(Y)}$$

f) Matrice des variances-covariances

$$M(X, Y) = \begin{pmatrix} V(x) & \text{Cov}\,(x, y) \\ \text{Cov}\,(x, y) & V(y) \end{pmatrix}$$

1.6 Indépendance de deux variables aléatoires *X* et *Y*

➤ Définition

Deux variables X et Y sont indépendantes si, quel que soit deux événements $X \in A$ et $Y \in B$, on a :

$$P(X \in A\,,\, Y \in B) = P(X \in A) \times P(Y \in B)$$

➤ Propriétés

Deux variables X et Y sont indépendantes si et seulement si :

▷ $F(x, y) = F(x, .) \times F(., y)$

▷ $f(x, y) = f(x, .) \times f(., y)$

▷ $f(x/y) = f(x, .) \qquad f(y/x) = f(., y)$

▷ $D_{(X,Y)} \equiv D_X \times D_Y$

➤ Variance de la somme

Si les deux variables X et Y sont indépendantes alors,

▷ $E(XY) = E(X) \times E(Y)$

▷ $\text{Cov}\,(X, Y) = 0$

▷ $V(X + Y) = V(X) + V(Y)$

1.7 Probabilités individuelles

a) Lois discrètes

Le tableau ci-après rappelle la définition, l'espérance et la variance des six lois discrètes les plus courantes

Loi	Définition	Espérance	Variance
Uniforme	$P(X=x)=\frac{1}{n} \quad x \in \{1,2,\ldots,n\}$	$\frac{n+1}{2}$	$\frac{n^2-1}{12}$
Bernoulli	$P(X=x)=p^x(1-p)^{1-x} \quad x \in \{0,1\}$	p	$p(1-p)$
Pascal	$P(X=x)=p(1-p)^{1-x} \quad x \in \{1,2,\ldots\}$	$\frac{1}{p}$	$\frac{q}{p^2}$
Binomiale	$P(X=x)=C_n^x p^x(1-p)^{n-x} \quad x \in \{0,1,\ldots n\}$	np	$np(1-p)$
Hyper-géométrique	$P(T=t)=\frac{C_{N_p}^t C_{N-N_p}^{n-t}}{C_N^n}$ $t \in \{\min(0,n-Np),\ldots \max(n,Np)\}$	np	$np(1-p)\frac{N-n}{N-1}$
Poisson	$P(X=x)=e^{-m}\frac{m^x}{x!} \quad x \in \{0,1,\ldots n\}$	m	m

b) Lois continues

De même, le tableau ci-dessous rappelle les propriétés de quelques lois continues :

Loi	Définition	Espérance	Variance
Uniforme sur $[a,b]$	$f(x)=\frac{1}{b-a} \quad x \in [a,b]$	$\frac{b-a}{2}$	$\frac{(b-a)^2}{12}$
Gauss $LG(m,\sigma)$	$f(x)=\frac{1}{\sigma\sqrt{2\pi}}\exp\left[-\frac{(x-m)^2}{2\sigma^2}\right] \quad x \in]-\infty,+\infty[$	m	σ^2
Exponentielle de paramètre λ	$f(x)=\lambda\exp(-\lambda x) \quad x \in [0,+\infty[$	$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$
Gamma $\gamma(\lambda,r)$	$f(x)=\frac{\lambda}{\Gamma(r)}e^{-\lambda x}(\lambda x)^{r-1}$	$\frac{r}{\lambda}$	$\frac{r}{\lambda^2}$

1.8 Lois de la somme de variables indépendantes connues

➤ Binomiale

$$B(n_1,p)+B(n_2,p)=B(n_1+n_2,p)$$

➤ Poisson

$$P(\lambda_1)+P(\lambda_2)=P(\lambda_1+\lambda_2)$$

➤ Normales

$$\sum_{i=1}^{n} a_i LG(m_i,\sigma_i)=LG\left(\sum_{i=1}^{n} a_i m_i, \sqrt{\sum_{i=1}^{n} a_i^2\sigma_i^2}\right)$$

ÉNONCÉS DES EXERCICES

1.1 *n étudiants sont soumis à un test d'aptitude. Pour coder ce test afin de le rendre anonyme, le responsable propose d'indiquer sur la fiche-test de chaque étudiant les quatre chiffres correspondant au jour et au mois de naissance de celui-ci. Quelle est la probabilité que deux étudiants aient le même code ?

1.2 *Après une marée noire en Bretagne, l'organisme de protection des oiseaux de mer a évalué à 20 000 la population de sternes au large du Finistère. 500 d'entre eux ont été bagués. Un an après, on capture 100 sternes dans cette zone. Calculer la probabilité :

1. de ne pas avoir d'oiseau bagué ?

2. d'avoir au moins deux sternes baguées ?

1.3 *Un dépistage systématique concernant un éventuel trouble de l'audition est effectué à la naissance. On sait que 2 % des nouveaux-nés présentent des troubles de l'audition. Ce dépistage commence par un test donnant 95 % de résultats positifs pour les nouveaux-nés atteints de ces troubles et 6 % de résultats positifs pour les bébés indemnes de ces troubles.

1. Quelle est la probabilité qu'un nouveau-né pris au hasard soit atteint de ces troubles sachant que le test a donné un résultat positif ?

2. Quelle est la probabilité qu'un nouveau-né pris au hasard soit indemne de ces troubles sachant que le test a donné un résultat négatif ?

1.4 *Dans une carrière de marbre, un contrôle est effectué sur des dalles destinées à la construction. La surface des dalles est vérifiée pour détecter d'éventuels éclats ou taches. Il a été constaté qu'en moyenne il ya 1,2 défaut par dalle et que le nombre de défauts par dalle suit une loi de Poisson.

1. Quel est le paramètre de cette loi de Poisson ? Quelles sont les valeurs possibles de la variable ?

2. Quelle est la probabilité d'observer plus de 2 défauts par dalle ?

3. L'entreprise présente à ses clients deux catégories de dalles : celles présentant moins de deux défauts (qualité ***) et celles présentant au moins deux défauts (qualité **). Quelle est la probabilité d'observer au moins deux défauts sur une dalle ? Quelle est alors la proportion de dalles de qualité ** ?

4. Sur 500 dalles contrôlées, quel est le nombre attendu ne présentant aucun défaut ?

1.5 ** Une grande mutuelle d'assurances envisage d'éventuels changements de tarifs. Pour cela, elle a étudié le risque d'accident automobile de ses assurés en fonction de l'ancienneté de leur permis. Parmi ses assurés, il y a 20 % de jeunes ayant leur permis depuis moins de 5 ans et le risque d'accident de ces jeunes conducteurs est de 0,4. Le risque d'accident des assurés ayant leur permis depuis plus de 5 ans est de 0,125.

1. Si on choisit au hasard 10 jeunes conducteurs, quelle est la probabilité d'en voir au moins un ayant un accident dans l'année ?

2. Même question avec 10 assurés ayant leur permis depuis plus de 5 ans.

3. Si on prend au hasard 10 assurés, quelle est la probabilité d'en voir au moins un ayant un accident dans l'année ?

1.6 ** Un fabricant d'ordinateurs portables souhaite vérifier que la période de garantie qu'il doit associer au disque dur correspond à un nombre pas trop important de retours de ce composant sous garantie. Des essais en laboratoire ont montré que la loi suivie par la durée de vie, en années, de ce composant est la loi exponentielle de moyenne 4.

1. Préciser la fonction de répartition de cette loi ainsi que son espérance $E(X)$ et son écart-type σ.

2. Quelle est la probabilité qu'un disque dur fonctionne sans défaillance plus de quatre ans ?

3. Quelle est la probabilité qu'un disque dur fonctionne sans défaillance six ans au moins, sachant qu'il a fonctionné déjà cinq ans.

4. Quelle est la probabilité que la durée de vie appartienne à l'intervalle : $[E(X) - \sigma, E(X) + \sigma]$?

5. Pendant combien de temps, 50 % des disques durs fonctionnent-ils sans défaillance ?

6. Donner la période de garantie optimum pour remplacer moins de 15 % des disques durs sous garantie.

1.7 ** On estime que 1 400 passagers ont réservé, le vendredi soir, sur le TGV Paris-Nantes de 19h30.

Les portes du train ouvrent une demi-heure avant le départ.

Parmi les usagers, 50 arrivent avant l'ouverture des portes et 70 arrivent trop tard.

On considère la variable aléatoire X, égale à la date d'arrivée d'un voyageur calculée par rapport à 19h30. ($X = 0$ à 19h30 et X est exprimé en minutes).

1. En admettant que cette variable X suit une loi $LG(m, \sigma)$, calculer m et σ.

2. Déterminer l'heure à laquelle les portes du train doivent être ouvertes pour qu'il n'y ait pas plus de 20 usagers qui attendent sur le quai.

3. Calculer le nombre de voyageurs ayant manqué le train si celui-ci accuse un retard de 5 minutes.

1.8 *** Le modèle suivant peut être utilisé pour représenter le nombre de blessés dans les accidents de la circulation au cours d'un week-end.

Le nombre d'accidents suit une loi de Poisson de paramètre λ.

Le nombre de blessés par accident, suit une loi de Poisson de paramètre μ.

Le nombre total de blessés est donc :

$$S = X_1 + X_2 + \cdots + X_N$$

S est la somme d'un nombre aléatoire de variables de Poisson, indépendantes et de même loi.

1. Donner une expression pour $P(S = s)$.

2. Calculer $P(S = 0)$.

3. Calculer $E(S)$ et $V(S)$.

1.9 *** Soit X une variable aléatoire suivant une loi de densité :

$$f(x) = \frac{1}{\sqrt{\pi x}} \times e^{-x} \quad \text{pour} \quad x > 0$$

Soit Y une autre variable aléatoire. On suppose que la loi conditionnelle de Y sachant X est une loi normale de paramètres $m = 0$ et $\sigma^2 = \dfrac{1}{2X}$

1. Calculer la loi du couple (Y, X)

2. Quelle est la loi conditionnelle de X sachant Y ?

3. En déduire $E\left[X/Y\right]$.

ÉNONCÉS DES PROBLÈMES

Problème 1.1

Un avion long-courrier peut transporter 100 passagers et leurs bagages. Les 100 places ont été réservées. Il pèse 120 tonnes sans passager ni bagages, mais l'équipage compris et le plein de carburant effectué. Les consignes de sécurité interdisent au commandant de décoller si le poids de l'appareil chargé dépasse 129,49 tonnes.

Le poids d'un passager suit une loi d'espérance mathématique 70 kg et d'écart-type 10 kg.

Le poids de ses bagages suit une loi d'espérance mathématique 20 kg et d'écart-type 10 kg.

Toutes ces variables aléatoires sont indépendantes.

1. Calculer l'espérance mathématique et l'écart-type du poids total de l'appareil au moment du décollage, tous les passagers et leurs bagages ayant été embarqués.

2. Le poids d'un voyageur et celui de ses bagages suivent des lois dont on ne connaît pas la nature. Par contre, l'espérance mathématique et la variance de chacune de ces lois ont les valeurs données précédemment. Calculer une limite supérieure de la probabilité pour que le commandant refuse d'embarquer une partie des bagages afin que le poids de l'appareil ne dépasse pas 129,42 tonnes.

3. On suppose en fait que le poids d'un voyageur suit une loi normale $LG(70\,;10)$ et celui des bagages suit une loi normale $LG(20\,;10)$.

Calculer la probabilité pour que le commandant refuse d'embarquer une partie des bagages afin que le poids de l'appareil ne dépasse pas 129,42 tonnes.

Expliquer la différence avec le résultat précédent.

Problème 1.2

On dispose de n variables aléatoires réelles $(X_1, \ldots, X_n)$ mutuellement indépendantes, de même loi, de fonction de répartition F de classe C^2 sur R. La densité de ces variables aléatoires est notée f.

Soit $(Y_1, \ldots, Y_n)$ la suite des X_i ordonnées de façon croissante.

Dans cet exercice, on s'intéresse à la loi du couple (Y_n, Y_1).

La fonction de répartition du couple (Y_n, Y_1) est définie par $\Phi(x, y) = P[(Y_n \leqslant x) \cap (Y_1 \leqslant y)]$.

1. Calculer $P[(Y_n \leqslant x) \cap (Y_1 > y)]$ et en déduire la fonction de répartition Φ du couple (Y_n, Y_1).

2. Déterminer la densité ϕ du couple (Y_n, Y_1) en fonction de F et de f.

3. Dans cette question, les variables X_i suivent toutes des lois uniformes sur l'intervalle [0,1].

a) Établir les lois de Y_n et de Y_1 et calculer leurs espérances mathématiques.

b) Calculer $E[(Y_n - Y_1)^2]$.

c) En déduire $E(Y_1 Y_n)$.

d) Calculer la covariance de Y_1 et Y_n ainsi que leur coefficient de corrélation linéaire r_{Y_n,Y_1}.

DU MAL À DÉMARRER ?

1.1 Il est plus simple, dans un premier temps, de chercher la probabilité de l'événement complémentaire.

1.2 La capture des oiseaux se faisant en globalité, on recherche un nombre de combinaisons sans répétition.

1.3 Dans ce type d'exercice, il est essentiel de mettre en évidence les événements en présence.

1.4 De la valeur moyenne de la variable, on peut déduire la valeur du paramètre de la loi.

1.5 Pour chacune des deux premières questions, on identifiera les paramètres de la loi suivie par la variable considérée. Dans la troisième question, la population totale des assurés est constituée du mélange des deux catégories étudiées ; il faut alors chercher la probabilité qu'un assuré quelconque ait un accident.

1.6 L'espérance de la loi exponentielle est l'inverse du paramètre.

1.7 Le nombre de passagers arrivés avant l'ouverture des portes et celui des passagers arrivés en retard permettent de calculer les paramètres de la loi.

1.8 La difficulté de l'exercice réside dans le fait que le nombre d'accidents, dans un week-end donné, est lui-même une réalisation d'une variable aléatoire. On est donc amené, dans un premier temps, à chercher une probabilité conditionnelle.

1.9 La connaissance de la loi conditionnelle de Y sachant X et de la loi marginale de X nous permet de déterminer la loi du couple.

Problème 1.1

On détermine très facilement l'espérance du poids total T de l'appareil et l'indépendance des variables permet de déterminer aussi la variance de T. Dans la question 2, ne connaissant pas la loi des variables, on peut utiliser l'inégalité de Bienaymé-Tchebichev. En revanche, dans la question suivante, on peut déterminer la valeur exacte de la probabilité recherchée.

Problème 1.1

Il est essentiel de remarque que dans la série ordonnée, Y_1 est le minimum et Y_n le maximum.

CORRIGÉS DES EXERCICES

1.1 Il y a autant de codage possibles que de n-uplets formé des n jours anniversaires, c'est à dire 365^n. Un codage formé de nombres tous différents correspond à un arrangement de n dates anniversaires prises parmi 365, c'est-à-dire : A_{365}^n. La probabilité d'avoir n codages tous différents est donc :

$$\frac{A_{365}^n}{365^n}$$

La probabilité d'avoir au moins deux codages similaires est donc :

$$p = 1 - \frac{A_{365}^n}{365^n}$$

Ainsi pour $n = 30$, on a déjà $p = 0{,}7$

Ce qu'il faut retenir de cet exercice

Ce résultat surprend toujours car on le confond souvent avec la probabilité qu'un étudiant ait le même jour de naissance qu'une personne donnée.

1.2 Le nombre de dispositions que l'on peut faire en choisissant 100 sternes parmi 20 000 est : $C_{20\,000}^{100}$. Les oiseaux non bagués sont au nombre de 19 500 donc le nombre de combinaisons de 100 sternes non baguées parmi 20 000 est : $C_{19\,500}^{100}$.

1. La probabilité de ne pas avoir d'oiseau bagué est alors le quotient :

$$p_0 = \frac{C_{19\,500}^{100}}{C_{20\,000}^{100}} \simeq 0,0790$$

2. La probabilité d'avoir exactement un oiseau bagué est (nombre de choix de l'oiseau bagué = 500) :

$$p_1 = \frac{500 \times C_{19\,500}^{99}}{C_{20\,000}^{100}} \simeq 0,2036$$

La probabilité d'avoir au moins deux sternes baguées est alors :

$$p = 1 - p_0 - p_1 \simeq 0,7174$$

Le calcul de ces probabilités est fastidieux. La population de ces oiseaux étant importante dans cette région, on peut considérer que chacun d'eux a la probabilité $p = 500/20\,000 = 1/40$ d'être bagué. La probabilité de ne pas avoir de sterne baguée est alors :

$$p_0 = \left(\frac{39}{40}\right)^{100} \simeq 0{,}0795$$

$$p_1 = C_{100}^{1}\left(\frac{1}{40}\right)\left(\frac{39}{40}\right)^{99} \simeq 0{,}2039$$

$\Rightarrow p = 0,7166$

Ce qu'il faut retenir de cet exercice

La remarque suggère l'approximation de la loi hypergéométrique par la loi binomiale, approximation justifiée par la petitesse du rapport entre la taille de l'échantillon et la taille de la population.

1.3 On va définir deux événements aléatoires :

A : le nouveau-né est atteint d'un trouble de l'audition. $P(A) = 0{,}02$

B : le test est positif. On ne connaît pas $P(B)$ mais on connaît les deux probabilités conditionnelles suivantes :

$$P(B/\overline{A}) = 0{,}06 \quad \text{et} \quad P(B/A) = 0{,}95$$

1. On cherche la probabilité qu'un nouveau-né pris au hasard soit atteint de ces troubles sachant que le test a donné un résultat positif c'est-à-dire : $P(A/B)$. Pour cela, on utilise la formule de Bayes :

$$P(A/B) = \frac{P(B/A) \times P(A)}{P(B)} = \frac{P(B/A) \times P(A)}{P(B/A)P(A) + P(B/\overline{A})P(\overline{A})}$$

$$P(A/B) = \frac{0{,}95 \times 0{,}02}{0{,}95 \times 0{,}02 + 0{,}06 \times 0{,}05} = 0{,}244\,2$$

2. De la même façon, la probabilité qu'un nouveau-né pris au hasard soit indemne de ces troubles sachant que le test a donné un résultat négatif est égale à :

$$P(\overline{A}/\overline{B}) = \frac{0{,}94 \times 0{,}98}{0{,}94 \times 0{,}98 + 0{,}05 \times 0{,}02} = 0{,}998\,9$$

Ce qu'il faut retenir de cet exercice

Ce type de raisonnement, très utilisé dans les travaux en Médecine, utilise la formule de Bayes pour trouver la probabilité conditionnelle recherchée.

1.4 **1.** Soit X la variable : « nombre de défauts par dalle ». La loi de X est une loi de Poisson. Son paramètre est égal à la moyenne observée sur l'échantillon : $\lambda = 1{,}2$. Les valeurs possibles de X sont les entiers positifs.

2. $P(X > 2) = 1 - P(X \leqslant 2)$.

Or

$$P(X = 0) = e^{-1{,}2}\,,\ P(X = 1) = 1{,}2 \times e^{-1{,}2}\,,\ P(X = 2) = \frac{1{,}2^2}{2!} \times e^{-1{,}2}$$

$$P(X = 0) = 0{,}301\,,\ P(X = 1) = 0{,}361\,,\ P(X = 2) = 0{,}217$$

$$P(X > 2) = 1 - P(X \leqslant 2) = 1 - 0{,}879 = 0{,}122$$

3. La probabilité d'observer au moins deux défauts sur une dalle est alors :

$$P(X \geqslant 2) = 1 - P(X \leqslant 1) = 1 - e^{-1{,}2} - 1{,}2 \times e^{-1{,}2} = 0{,}338$$

La proportion de dalles de qualité ** est donc 33,8 % :

4. Sur les 500 dalles contrôlées, le nombre attendu ne présentant aucun défaut est :

$$500 \times P(X = 0) \approx 150$$

Ce qu'il faut retenir de cet exercice

La loi de Poisson caractérise fréquemment les événements rares.

1.5 **1.** Soit X la variable aléatoire : nombre de jeunes conducteurs accidentés parmi les 10 choisis. Chacun de ces jeunes a une probabilité de 0,4 d'avoir un accident dans l'année et ceci indépendamment les uns des autres. La loi de la variable X est donc la loi binomiale de paramètres $n = 10$ et $p = 0,4$.

$$\Rightarrow \quad P(X \geqslant 1) = 1 - P(X = 0) = 1 - 0,6^{10} = 0,994$$

2. En raisonnant de la même façon avec Y la variable aléatoire : nombre de conducteurs de plus de 5 ans de permis, accidentés, parmi les 10 choisis, cette variable suit une loi binomiale de paramètres $n = 10$ et $p = 0,125$.

$$\Rightarrow \quad P(Y \geqslant 1) = 1 - P(Y = 0) = 1 - 0,875^{10} = 0,737$$

3. Cette fois, on s'intéresse à la variable Z nombre d'assurés accidentés parmi 10 choisis au hasard dans la population des assurés. Il nous faut chercher la probabilité $p*$ qu'un tel assuré ait un accident. Si on considère les événements : J : être jeune conducteur et A : avoir un accident dans l'année, on connaît les probabilités conditionnelles :

$$P(A/J) = 0,4 \quad , \quad P(A/\overline{J}) = 0,125$$

ainsi que $P(J) = 0,2$. Le théorème des probabilités totales donne alors :
$p* = 0,2 \times 0,4 + 0,8 \times 0,125 = 0,18$. On en déduit finalement :

$$P(Z \geqslant 1) = 1 - P(Z = 0) = 1 - 0,82^{10} = 0,86$$

Ce qu'il faut retenir de cet exercice

Dans la troisième question, on a un mélange de deux populations et le théorème des probabilités totales (ou formule de décomposition) permet de déterminer la probabilité d'accident.

1.6 **1.** La loi suivie par la durée de vie, en années, de ce composant est la loi exponentielle de moyenne 4. Sa densité est :

$$f(x) = 0,25e^{-0,25x} \quad \text{pour} \quad x \geqslant 0$$

La durée de vie moyenne est égale à $1/0,25 = 4$ ans et son écart-type est $\sigma = 4$. La fonction de répartition est :

$$F(x) = 0 \quad \text{si} \quad x < 0$$

$$F(x) = \int_0^x f(t)dt = 1 - e^{-0,25x} \quad \text{si} \quad x \geqslant 0$$

2. $P(X > 4) = 1 - F(4) = \exp(-1) = 0,368$

3. Par définition d'une probabilité conditionnelle :

$$P(X \geqslant 6/X > 5) = \frac{P(X \geqslant 6)}{P(X > 5)} = e^{-0,25(6-5)} = e^{-0,25} = 0,78$$

Il s'agit d'un phénomène sans mémoire.

4. $P[E(X) - \sigma < X < E(X) + \sigma] = P(0 < X < 8) = F(8) = 0,865$

5. On cherche la durée d durant laquelle 50 % des disques durs fonctionnent sans défaillance : $P(X > d) = 1 - F(d) = 0,5$. D'où $\exp(-0,25d) = 0,5$. On obtient $d = 2,77$ ans.

6. On cherche la durée t telle que : $P(X < t) \leqslant 0,15$. D'où :

$$P(X \geqslant t) = \exp(-0,25t) = 0,85 \quad \Rightarrow \quad t = -\frac{\ln 0,85}{0,25} \simeq 0,61$$

On pourra prendre une période de garantie de 7 mois.

Ce qu'il faut retenir de cet exercice

Le résultat de la troisième question est une caractéristique de la loi exponentielle. La probabilité que le matériel dure une année de plus ne dépend pas de l'instant initial.

1.7 **1.** L'origine des heures d'arrivée est placée à 19h30. X est l'heure d'arrivée d'un voyageur comptée à partir de 19h30 et exprimée en minutes.

Parmi l'ensemble des 1 400 passagers, X, variable statistique, représente l'heure d'arrivée des voyageurs, qui est décrite par sa distribution statistique.

Si on tire un voyageur au hasard , X est la variable aléatoire « heure d'arrivée du voyageur », dont la loi-parente est la loi de description statistique précédente et dont les probabilités peuvent être calculées à partir des fréquences observées dans l'ensemble de la population.

La variable aléatoire X suit la loi $LG(m, \sigma)$. Soit U la variable centrée réduite associée. On sait que 30 passagers arrivent une demi-heure avant le départ :

$$P(X \leqslant -30) = P\left(\frac{X - m}{\sigma} \leqslant \frac{-30 - m}{\sigma}\right) = P(U \leqslant u_0) = \frac{50}{1\,400}$$

$$P(U \leqslant u_0) = 1 - P(U \leqslant -u_0) = 0,035\,7 \qquad P(U \leqslant -u_0) = 0,964\,3$$

Dans la table de la loi normale centrée réduite, on lit :

$$F(1,81) = 0,964\,9 \quad \text{et} \quad F(1,80) = 0,964\,1$$

Par interpolation linéaire on a donc :

$$\frac{-u_0 - 1{,}80}{1{,}81 - 1{,}80} = \frac{0{,}964\,3 - 0{,}964\,1}{0{,}964\,9 - 0{,}964\,1} = \frac{0{,}000\,2}{0{,}000\,8} = \frac{1}{4}$$

$$\text{D'où :} \quad u_0 = -1{,}803 \quad \text{et} \quad -u_0 = 1{,}803 = \frac{30+m}{\sigma}$$

Par ailleurs, 70 passagers arrivent trop tard :

$$P(X \geqslant 0) = P\left(\frac{X-m}{\sigma} \geqslant \frac{-m}{\sigma}\right) = P\left(U \geqslant \frac{-m}{\sigma}\right) = \frac{70}{1\,400}$$

$$P(U \geqslant u_1) = 1 - P(U \leqslant u_1) = 0{,}05 \quad \text{soit} \quad P(U \leqslant u_1) = 0{,}95$$

Dans la table de la loi normale centrée réduite on lit :

$$F(1{,}64) = 0{,}949\,5 \quad \text{et} \quad F(1{,}65) = 0{,}950\,5$$

Par interpolation linéaire on a donc :

$$u_1 = 1{,}645 = -\frac{m}{\sigma}$$

On a donc le système :

$$\left\{\frac{30+m}{\sigma} = 1{,}803\,; \; -\frac{m}{\sigma} = 1{,}645\right\}$$

$$\Rightarrow \quad \{m = -14{,}31 \text{ mn} \quad ; \quad \sigma = 8{,}70 \text{ mn}\}$$

2. La variable aléatoire X suit la loi $LG(-14{,}31; 8{,}7)$

On doit déterminer a pour que :

$$P(X \leqslant a) = P\left(\frac{X+14{,}31}{8{,}7} \leqslant \frac{a+14{,}31}{8{,}7}\right)$$

$$P(X \leqslant a) = P\left(U \leqslant u_3 = \frac{a+14{,}31}{8{,}7}\right) = \frac{20}{1\,400} = 0{,}014\,3$$

$$P(U \leqslant u_3) = P(U \geqslant -u_3) = 1 - P(U \leqslant -u_3) = 0{,}014\,3$$

$$P(U \leqslant -u_3) = 1 - 0{,}014\,3 = 0{,}985\,7$$

Dans la table de la loi normale centrée réduite on lit : $F(2{,}19) = 0{,}985\,7$

$$-u_3 = \frac{-14{,}31 - a}{8{,}7} = 2{,}19 \quad \text{et} \quad a = -33{,}36 \text{ mn} = 33 \text{ mn et } 20 \text{ s}$$

Il faut donc ouvrir les portes du train **33 minutes** avant le départ.

3. Soit n le nombre de personnes ayant manqué le train, parti avec 5 mn de retard :

$$P(X \geqslant 5) = p = \frac{n}{1\,400}$$

$$P(X \geqslant 5) = P\left(\frac{X + 14{,}31}{8{,}7} \geqslant \frac{5 + 14{,}31}{8{,}7}\right) = P\left(U \geqslant \frac{19{,}31}{8{,}7}\right)$$

$$P(X \geqslant 5) = P(U \geqslant 2{,}217)$$

$$P(U \geqslant 2{,}21) = 1 - P(U \leqslant 2{,}21) = 1 - F(2{,}21) = 1 - 0{,}986\,4 = 0{,}013\,6$$

$$P(U \geqslant 2{,}22) = 1 - P(U \leqslant 2{,}22) = 1 - F(2{,}22) = 1 - 0{,}986\,8 = 0{,}013\,2$$

Par interpolation linéaire on a donc (en utilisant la taille $LG(0, 1)$) :

$$\frac{p - 0{,}013\,6}{0{,}013\,2 - 0{,}013\,6} = \frac{2{,}217 - 2{,}21}{2{,}22 - 2{,}21} = \frac{0{,}007}{0{,}01} = \frac{7}{10}$$

$$p = 0{,}013\,32 \quad \text{et} \quad n = 1\,400 \times p = 18{,}648$$

Il y aura donc, si le train accuse 5 **minutes** de retard, 19 **personnes** qui manqueront le train.

Ce qu'il faut retenir de cet exercice

Seule la loi normale centrée réduite est tabulée. Aussi, dès qu'on a une variable X suivant une loi normale de paramètres m et σ, on introduit la variable centrée réduite associée : $U = \dfrac{(X - m)}{\sigma}$.

1.8 Soit N la variable aléatoire représentant le nombre d'accidents. La loi de N est la loi de Poisson de paramètre λ.

Soit X la variable aléatoire représentant le nombre de blessés par accidents. La loi de X est la loi de Poisson de paramètre μ.

Soit S la variable aléatoire représentant le nombre total de blessés :

$$S = \sum_{i=1}^{N} X_i$$

S est donc la somme d'un nombre aléatoire N de variables de Poisson, indépendantes, suivant toutes la même loi.

1. Si on connaît le nombre d'accidents du week-end, on peut alors connaître le nombre de blessés dans le week-end en utilisant la somme de variables de Poisson : on va donc utiliser la loi de probabilité conditionnelle :

$$P(S = s/N = n) = \frac{e^{-\mu n}(\mu n)^s}{s!} \quad \Rightarrow \quad P(S = s) = \sum_{n=0}^{\infty} \frac{e^{-\mu n}(\mu n)^s}{s!} \frac{e^{-\lambda}\lambda^n}{n!}$$

$$P(S = s) = \frac{e^{-\lambda}\mu^s}{s!} \sum_{n=0}^{\infty} \frac{\lambda^n n^s e^{-\mu n}}{n!}$$

2.

$$P(S = 0) = e^{-\lambda} \sum_{n=0}^{\infty} \frac{\lambda^n e^{-\mu n}}{n!} = \exp[-\lambda(1 - e^{-\mu})]$$

3. $E(S) = E[E(S/N)]$

$E(S/N = n) = n\mu \quad \Rightarrow \quad E(S/N) = N\mu$

$E(S) = E(N\mu) = \mu E(N) = \mu\lambda$

$V(S) = E[V(S/N)] + V[E(S/N)]$

$E(S/N) = N\mu$ et $V(S/N) = N\mu$

$E[V(S/N)] = E(N\mu) = \mu E(N) = \mu\lambda$

$V[E(S/N)] = V(N\mu) = \mu^2 V(N) = \mu^2\lambda$

$\Rightarrow \quad V(S) = \mu\lambda + \mu^2\lambda = \mu\lambda(1 + \mu)$

Ce qu'il faut retenir de cet exercice

L'application des théorèmes de l'espérance totale et de la variance totale trouve ici tout son sens.

1.9 **1.** Par définition, la densité conditionnelle de Y sachant X est égale à :

$$f(y/x) = \frac{f(y, x)}{f(x)}$$

On en déduit la loi du couple, pour $x > 0$:

$$f(y, x) = \frac{1}{\sqrt{\pi x}} \times e^{-x} \times \frac{1}{\sqrt{2\pi/2x}} \times e^{-\frac{1}{2}\times 2xy^2}$$

$$f(y, x) = \frac{1}{\pi} \times e^{-x(1+y^2)}$$